Predictive Coarse-Graining

M. Schöberl[1], N. Zabaras[2,3], P.S. Koutsourelakis[1]

[1]Continuum Mechanics Group
Technische Universität München

[2]Institute for Advanced Study
Technische Universität München

[3]Warwick Centre of Predictive Modeling
University of Warwick

November 27, 2015

Motivation Coarse-Graining

Atomistic simulation for obtaining insights of chemical and physical process of complex systems.

Difficulty

- Complex interactions
- Long-range interactions
- Small time- and length-scales

→ Exceeding computational tractability

A coarse description allows us

- to evaluate larger systems during larger time intervals
- to gain understanding of physics of the system.

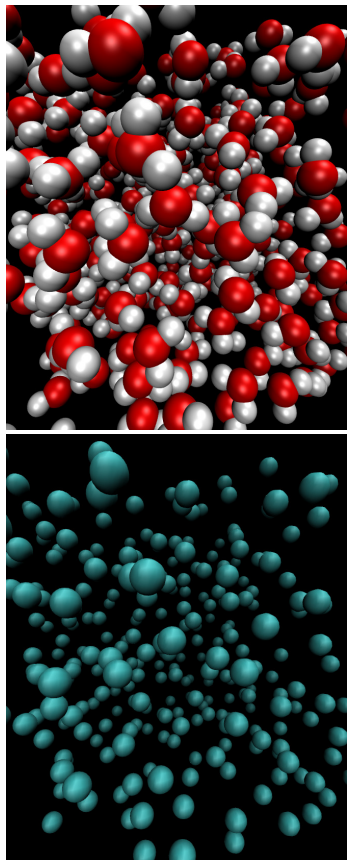


Figure :
Coarse-graining water

Motivation

Approach

- Describing system with less degrees of freedom
- Determining optimal parameter set for given parametrization of a coarse-grained potential
- Leading to point estimates of its parametrization and thus also in predictions
- Predictions performed on coarse scale

How can we quantify the uncertainty induced by a coarse description and the loss of information?

How can we reconstruct fine configurations given a coarse configuration?

General coarse-graining problem

Fine-scale degrees of freedom $\mathbf{x} \in \mathcal{M}$ with $\mathcal{M} \subset \mathbb{R}^n$, $n \gg 1$ in equilibrium described by a Boltzmann-type PDF:

Fine-scale description

$$p_{\mathrm{f}}(\mathbf{x}|\beta) = \frac{\exp\{-\beta U(\mathbf{x})\}}{Z(\beta)}$$

- Potential $U(\mathbf{x})$
- Inverse Temperature $\beta = \frac{1}{k_b T}$, with temperature T
- Partition function $Z(\beta) = \int_{\mathcal{M}} \exp\{-\beta U(\mathbf{x})\}\, d\mathbf{x}$

General coarse-graining problem

Coarse-scale description

$$p_c(\mathbf{X}|\beta) = \frac{\exp\{-\beta U_c(\mathbf{X}, \boldsymbol{\theta}_c)\}}{Z_c(\boldsymbol{\theta}_c, \beta)}$$

- Coarse variables $\mathbf{X} \in \mathcal{M}_c$, $\mathcal{M}_c \subset \mathbb{R}^{n_c}$, $n_c \ll n$
- Coarse-grained potential selected $\hat{U}_c$ selected out of candidate potentials $\hat{U}_c \in \mathcal{U}_c$
- Parametrization and shape selected by assuming $U_c(\mathbf{X}, \boldsymbol{\theta}_c)$ but any potential possible

How are fine and coarse configurations connected?

Connecting fine variables $\mathbf{x}$ with coarse variables $\mathbf{X}$ with coarse-graining map $\xi : \mathcal{M} \longrightarrow \mathcal{M}_c$

$$\mathbf{X} = \xi(\mathbf{x})$$

How to determine the effective coarse potential U_c?

Analytical solution for the effective coarse-grained potential:

$$\boxed{U_c(\mathbf{X}) = -\beta^{-1} \ln \int_{\mathcal{M}} \exp\{-\beta U(\mathbf{x})\}\, \delta(\boldsymbol{\xi}(\mathbf{x}) - \mathbf{X}) d\mathbf{x}}$$

Potential of mean force with respect to coarse-grained coordinates $\boldsymbol{\xi}(\mathbf{x})$. $\rightarrow$ intractable!

Numerical coarse-graining strategies:

- iterative Boltzmann inversion [5]
- inverse Monte Carlo [8]
- force matching [6]
- variational mean-field theory [9]
- relative entropy [10]

The relative entropy method[10]

$\rightarrow$ minimize a distance metric between $p_f(\mathbf{X}|\beta)$ and $p_c(\mathbf{X}|\beta)$ with the PDF $p_f(\mathbf{X}|\beta)$ sampling configuration $\mathbf{x}$ that maps to configuration $\mathbf{X}$

$$p_f(\mathbf{X}|\beta) \propto \int_{\mathcal{M}} p_f(\mathbf{x}|\beta)\delta(\boldsymbol{\xi}(\mathbf{x}) - \mathbf{X})d\mathbf{x}$$

Relative entropy method[10, 4, 2] (or KL-divergence)

$$\mathrm{KL}(p_f||p_c) = \int_{\mathcal{M}_c} p_f(\mathbf{X}|\beta) \ln\left(\frac{p_f(\mathbf{X}|\beta)}{p_c(\mathbf{X}|\beta)}\right) d\mathbf{X}$$

Formulation in $\mathcal{M}$:

$$\mathrm{KL}(p_f||p_c) = \int_{\mathcal{M}} p_f(\mathbf{x}|\beta) \ln\left(\frac{p_f(\mathbf{x}|\beta)}{p_c(\boldsymbol{\xi}(\mathbf{x})|\beta)}\right) d\mathbf{x} + \mathcal{S}_{\mathrm{map}}(\boldsymbol{\xi}(\mathbf{x}))$$

- with $\mathcal{S}_{\mathrm{map}} = \left\langle \int_{\mathcal{M}} \delta(\boldsymbol{\xi}(\mathbf{x}) - \mathbf{X})d\mathbf{x} \right\rangle_{p_f(\mathbf{x}|\beta)}$ measures the information loss induced by mapping $\boldsymbol{\xi}(\mathbf{x})$.

Relative entropy method

$\mathcal{S}_{\text{map}}(\boldsymbol{\xi}(\mathbf{x}))$ is independent of $\hat{U}_{\text{c}}$, the minimization of the KL-divergence is equivalent to minimization of

$$\mathcal{F}(\hat{U}_{\text{c}}) = \left\langle \ln \frac{p_{\text{f}}(\mathbf{x}|\beta)}{p_{\text{c}}(\boldsymbol{\xi}(\mathbf{x})|\beta)} \right\rangle_{p_{\text{f}}(\mathbf{x}|\beta)}$$

For any given m-dimensional parametrization $\boldsymbol{\theta}_{\text{c}} \in \boldsymbol{\Theta}_{\text{c}} \subset \mathbb{R}^m$ of the potential, the optimization problem follows:

$$\boxed{\boldsymbol{\theta}_{\text{c}}^* = \arg \min_{\boldsymbol{\theta}_{\text{c}} \in \boldsymbol{\Theta}_{\text{c}}} \mathcal{F}(U_{\text{c}}(\mathbf{X}, \boldsymbol{\theta}_{\text{c}}))}$$

Relative entropy method

Model Restrictions

- Relative entropy method has no capability to predict fine-scale configurations (no mapping coarse to fine)
  - Eg. for predicting correlations within fine configurations
- Predicting quantities of interest (QoI) by evaluating on fine-scale

$$\langle f(x) \rangle_{p_f(\mathbf{x}|\beta)} = \int_{\mathcal{M}} f(x) p_f(\mathbf{x}|\beta) d\mathbf{x}$$

- With coarse-grained description $p_c(\mathbf{X}|\beta)$ predictions in $\mathcal{M}_c$ but not in $\mathcal{M}$
  Difficulty:
  - How to define observable $g(\mathbf{X})$ for desired QoI in $\mathcal{M}_c$?
  - Are you able to predict same QoI with $\langle g(\mathbf{X}) \rangle_{p_c(\mathbf{X}|\beta)} = \int_{\mathcal{M}_c} g(\mathbf{X}) p_c(\mathbf{X}|\beta) d\mathbf{X}$
- No quantification of epistemic uncertainty: determine $\mathcal{S}_{\text{map}}(\boldsymbol{\xi}(\mathbf{x}))$

Proposed Model

Framework following **generative probabilistic models** and builds upon:

1. Coarse-scale PDF $p_c(\mathbf{X}|\boldsymbol{\theta}_c, \beta) \propto \exp\{-\beta U_c(\mathbf{X}, \boldsymbol{\theta}_c)\}$
   - Describing statistics of coarse variables $\mathbf{X}$
   - Parametrized by $\boldsymbol{\theta}_c$, no restrictions in shape of p_c

2. Probabilistic mapping from **coarse to fine** $p_{cf}(\mathbf{x}|\mathbf{X}, \boldsymbol{\theta}_{cf})$
   - Conditional PDF of $\mathbf{x}$ given the coarse variables $\mathbf{X}$
   - Parametrization of the probabilistic mapping by $\boldsymbol{\theta}_{cf} \in \boldsymbol{\Theta}_{cf} \subset \mathbb{R}^{m_{cf}}$
   - Deterministic mapping $\boldsymbol{\xi}(\mathbf{x})$ is not invertible (many to one map)

     $\rightarrow$ Probabilistic relation necessary
   - $\mathcal{S}_{map}$ not fixed for per definition. $\rightarrow$ Optimization with respect to $\boldsymbol{\theta}_{cf}$ possible.

3. Prior PDF for model parametrization
   - $p(\boldsymbol{\theta}_c)$
   - $p(\boldsymbol{\theta}_{cf})$

Proposed Model

- Proposed model describes a **Bayesian perspective** of coarse-graining equilibrium atomistic ensembles
- Model is data driven by N data points $\mathbf{x}^{(i)}$ in the fine-scale, denoted as $\mathbf{x}^{(1:N)} = \{\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(N)}\}$

Given the data $\mathbf{x}^{(1:N)}$ a posterior distribution on the model parameters $\boldsymbol{\theta}_{\text{cf}}$, $\boldsymbol{\theta}_{\text{c}}$ and the **latent** variables $\mathbf{X}^{(1:N)}$ can be defined.

- Coarse variables seen as **latent**
- To each observable $\mathbf{x}^{(i)}$ one latent Variable $\mathbf{X}^{(i)}$ is assigned
- $\mathbf{X}^{(i)}$ represents pre-image of the fine-scale observation $\mathbf{x}^{(i)}$

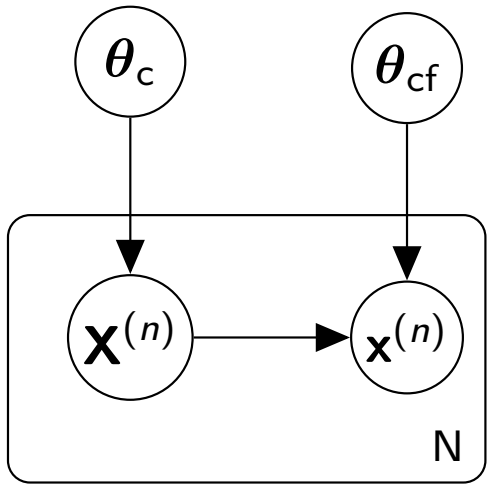
Bayesian coarse-graining - posterior

Posterior distribution

$$p(\boldsymbol{\theta}_{\mathrm{cf}}, \boldsymbol{\theta}_{\mathrm{cf}}, \mathbf{X}^{(1:N)} | \mathbf{x}^{(1:N)}) \propto p(\mathbf{x}^{(1:N)} | \boldsymbol{\theta}_{\mathrm{cf}}, \boldsymbol{\theta}_{\mathrm{c}}, \mathbf{X}^{(1:N)}) p(\boldsymbol{\theta}_{\mathrm{cf}}, \boldsymbol{\theta}_{\mathrm{c}}, \mathbf{X}^{(1:N)})$$
$$= \left(\prod_{i=1}^{N} p_{\mathrm{cf}}(\mathbf{x}^{(i)} | \boldsymbol{\theta}_{\mathrm{cf}}, \mathbf{X}^{(i)}) p_{\mathrm{c}}(\mathbf{X}^{(i)} | \boldsymbol{\theta}_{\mathrm{c}}) \right) p(\boldsymbol{\theta}_{\mathrm{cf}}) p(\boldsymbol{\theta}_{\mathrm{c}})$$

Use given data $\mathbf{x}^{(1:N)}$ for inferring the the posterior distribution or for determining MAP estimates of the model parameters $\boldsymbol{\theta}^*_{\mathrm{cf}}$, $\boldsymbol{\theta}^*_{\mathrm{c}}$.

- Bayesian formulation answers the questions of
  - model validation and
  - prediction



Bayesian coarse-graining

The posterior distribution leads to the predictive distribution for fine-scale configurations $\mathbf{x}$:

Predictive distribution

$$p(\mathbf{x}|\mathbf{x}^{(1:N)}) = \int p_{\mathrm{cf}}(\mathbf{x}|\mathbf{X}, \theta_{\mathrm{cf}}) p_{\mathrm{c}}(\mathbf{X}|\theta_{\mathrm{c}}) p(\theta_{\mathrm{c}}, \theta_{\mathrm{cf}}|\mathbf{x}^{(1:N)}) d\mathbf{X} d\theta_{\mathrm{cf}} d\theta_{\mathrm{cf}}$$

Simulate fine-scale system

1. Draw sample θ_{c}^*, $\theta_{\mathrm{cf}}^* \sim p(\theta_{\mathrm{c}}, \theta_{\mathrm{cf}}|\mathbf{x}^{(1:N)})$
2. Draw sample of coarse-scale description $\mathbf{X}^* \sim p_{\mathrm{c}}(\mathbf{X}|\theta_{\mathrm{c}}^*)$
3. Draw sample of fine-scale description $\mathbf{x}^* \sim p_{\mathrm{cf}}(\mathbf{x}^*|\mathbf{X}^*, \theta_{\mathrm{cf}}^*)$

Bayesian coarse-graining: ensemble averages

Approximating ensemble averages by

$$
\begin{aligned}
\langle f(\mathbf{x}) \rangle_{p_{\mathrm{f}}(\mathbf{x})} &= \int_{\mathcal{M}} f(\mathbf{x}) p_{\mathrm{f}}(\mathbf{x}) d\mathbf{x} \\
&= E[f(\mathbf{x})] \approx E[f(\mathbf{x})|\mathbf{x}^{(1:N)}] \\
&= \int_{\mathcal{M}} f(\mathbf{x}) p(\mathbf{x}|\mathbf{x}^{(1:N)}) d\mathbf{x} \\
&= \int f(\mathbf{x}) p_{\mathrm{cf}}(\mathbf{x}|\mathbf{X}, \boldsymbol{\theta}_{\mathrm{cf}}) p_{\mathrm{c}}(\mathbf{X}|\boldsymbol{\theta}_{\mathrm{c}}) p(\boldsymbol{\theta}_{\mathrm{c}}, \boldsymbol{\theta}_{\mathrm{cf}}|\mathbf{x}^{(1:N)}) d\mathbf{x} d\mathbf{X} d\boldsymbol{\theta}_{\mathrm{c}} d\boldsymbol{\theta}_{\mathrm{cf}}
\end{aligned}
$$

- Error arising from discrepancy between $E[f(\mathbf{x})]$ and $E[f(\mathbf{x})|\mathbf{x}^{(1:N)}]$

Bayesian coarse-graining: coarse-scale model p_c

Coarse-scale model

$$p_c(\mathbf{X}|\boldsymbol{\theta}_{cf}) = \frac{\exp\{-\beta U_c(\mathbf{X}, \boldsymbol{\theta}_c)\}}{Z_c(\boldsymbol{\theta}_c)}$$

Requirements

- Allow sufficient flexibility of $p_c(\mathbf{X})$
- U_c built from high-order interactions affording flexibility in $p_c(\mathbf{X})$

Bayesian coarse-graining: reconstruction map p_{cf}

Map from coarse to fine: $p_{\mathrm{cf}}(\mathbf{x}|\mathbf{X}, \boldsymbol{\theta}_{\mathrm{cf}})$

- Several fine-configurations map to same coarse-configuration
  - Probabilistic relation between a given coarse-configuration and a fine-configuration takes account of it. [7]
- Fast reconstruction ([7, 1]) of fine-scale configurations $\mathbf{x}$ given a coarse-configuration $\mathbf{X}$ desired.

Influence of p_{cf}

- **What is expected from the coarse variables $\mathbf{X}$ in terms of predicting the given data $\mathbf{x}^{(1:N)}$?**
- Adjusting $p_{\mathrm{c}}(\mathbf{X}|\boldsymbol{\theta}_{\mathrm{c}})$ so that $\mathbf{X}$ agrees best with data $\mathbf{x}^{(1:N)}$ connected with probabilistic mapping p_{cf}.

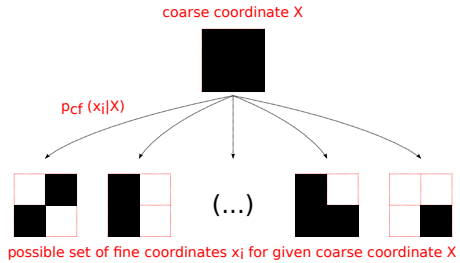


Figure : Configurations $\mathbf{x}$ given $\mathbf{X}$

Optimization

For learning the model parameters $\boldsymbol{\theta}_\mathrm{c}$ and $\boldsymbol{\theta}_\mathrm{cf}$ the expectation maximization algorithm is applied.

- Point estimates for maximizing the likelihood or posterior distribution are obtained.
- Approximation of posterior on $\boldsymbol{\theta}_\mathrm{cf}$ by Laplace (validated by full posterior sampling)

MAP estimate

$$\begin{aligned}\boldsymbol{\theta}^{MAP} &= \arg\max_{\boldsymbol{\theta}} \left[\log p(\boldsymbol{\theta}_\mathrm{c}, \boldsymbol{\theta}_\mathrm{cf}, \mathbf{x}^{(1:N)})\right] \\ &= \arg\max_{\boldsymbol{\theta}} \left[\log p(\mathbf{x}^{(1:N)}|\boldsymbol{\theta}_\mathrm{cf}, \boldsymbol{\theta}_\mathrm{cf}) + \log p(\boldsymbol{\theta}_\mathrm{c}) + \log p(\boldsymbol{\theta}_\mathrm{cf})\right]\end{aligned}$$

- In the following the ML-estimate is shown and will be extended by prior distributions $p(\boldsymbol{\theta})$ in a next step by adding log-priors $\log p(\boldsymbol{\theta}_\mathrm{c}) + \log p(\boldsymbol{\theta}_\mathrm{cf})$.

Optimization

- To obtain point estimates with MLE for the model parameters θ_c and θ_{cf} we maximize the log-likelihood by using the EM algorithm.
- The likelihood is defined as

$$\begin{aligned} p(\mathbf{x}^{(1:N)}|\boldsymbol{\theta}_c, \boldsymbol{\theta}_{cf}) &= \int p(\mathbf{x}^{(1:N)}, \mathbf{X}^{(1:N)}|\boldsymbol{\theta}_c, \boldsymbol{\theta}_{cf}) d\mathbf{X}^{(1:N)} \\ &= \int p_{cf}(\mathbf{x}^{(1:N)}|\mathbf{X}^{(1:N)}, \boldsymbol{\theta}_{cf}) p_c(\mathbf{X}^{(1:N)}|\boldsymbol{\theta}_c) d\mathbf{X}^{(1:N)} \\ &= \int \prod_{i=1}^{N} p_{cf}(\mathbf{x}^{(i)}|\mathbf{X}^{(i)}, \boldsymbol{\theta}_{cf}) p_c(\mathbf{X}^{(i)}|\boldsymbol{\theta}_c) d\mathbf{X}^{(1:N)} \end{aligned}$$

- augmenting the log-likelihood with an arbitrary density $q(\mathbf{X}^{(1:N)})$

$$\begin{aligned} \log p(\mathbf{x}^{(1:N)}|\boldsymbol{\theta}_c, \boldsymbol{\theta}_{cf}) &= \log \int q(\mathbf{X}^{(1:N)}) \frac{p_{cf}(\mathbf{x}^{(1:N)}|\mathbf{X}^{(1:N)}, \boldsymbol{\theta}_{cf}) p_c(\mathbf{X}^{(1:N)}|\boldsymbol{\theta}_c)}{q(\mathbf{X}^{(1:N)})} d\mathbf{X}^{(1:N)} \\ &\geq \int q(\mathbf{X}^{(1:N)}) \log \frac{p_{cf}(\mathbf{x}^{(1:N)}|\mathbf{X}^{(1:N)}, \boldsymbol{\theta}_{cf}) p_c(\mathbf{X}^{(1:N)}|\boldsymbol{\theta}_c)}{q(\mathbf{X}^{(1:N)})} d\mathbf{X}^{(1:N)} \\ &= \mathcal{L}(q(\mathbf{X}^{(1:N)}), \boldsymbol{\theta}_c, \boldsymbol{\theta}_{cf}) \end{aligned}$$

Expectation maximization

It holds the following decomposition for the lower bound:

$$\begin{aligned}\mathcal{L}(q(\mathbf{X}^{(1:N)}), \boldsymbol{\theta}_{\mathrm{c}}, \boldsymbol{\theta}_{\mathrm{cf}}) &= \int q(\mathbf{X}^{(1:N)}) \log \frac{p(\mathbf{x}^{(1:N)}|\mathbf{X}^{(1:N)}, \boldsymbol{\theta}_{\mathrm{cf}}) p(\mathbf{X}^{(1:N)}|\boldsymbol{\theta}_{\mathrm{c}})}{q(\mathbf{X}^{(1:N)})} d\mathbf{X} \\ &= -\mathrm{KL}(q(\mathbf{X}^{(1:N)})||p(\mathbf{X}^{(1:N)}|\mathbf{x}^{(1:N)}, \boldsymbol{\theta}_{\mathrm{cf}}, \boldsymbol{\theta}_{\mathrm{cf}})) + \log p(\mathbf{x}^{(1:N)})\end{aligned}$$

Decomposition of the log likelihood

$$\log p(\mathbf{x}^{(1:N)}|\boldsymbol{\theta}_{\mathrm{c}}, \boldsymbol{\theta}_{\mathrm{cf}}) = \mathcal{L}(q(\mathbf{X}^{(1:N)}), \boldsymbol{\theta}_{\mathrm{c}}, \boldsymbol{\theta}_{\mathrm{cf}}) + KL(q(\mathbf{X}^{(1:N)})||p(\mathbf{X}^{(1:N)}|\mathbf{x}^{(1:N)}, \boldsymbol{\theta}_{\mathrm{c}}, \boldsymbol{\theta}_{\mathrm{cf}})).$$

Maximizing $\mathcal{L}(q(\mathbf{X}^{(1:N)}), \boldsymbol{\theta}_{\mathrm{c}}, \boldsymbol{\theta}_{\mathrm{cf}})$ is equal to minimizing $\mathrm{KL}(q(\mathbf{X}^{(1:N)})||p(\mathbf{X}^{(1:N)}|\mathbf{x}^{(1:N)}, \boldsymbol{\theta}_{\mathrm{cf}}, \boldsymbol{\theta}_{\mathrm{cf}}))$

$p(\mathbf{X}^{(1:N)}|\mathbf{x}^{(1:N)}, \boldsymbol{\theta}_{\mathrm{cf}}, \boldsymbol{\theta}_{\mathrm{cf}})$ PDF over pre-images for $\mathbf{X}$ for the given data $\mathbf{x}^{(1:N)}$ and parameters $\boldsymbol{\theta}$.

Expectation maximization

For given parameters $\boldsymbol{\theta}_{\mathrm{cf}}$, $\boldsymbol{\theta}_{\mathrm{c}}$ the lower bound is maximized if we choose $q(\mathbf{X}^{(1:N)}) \propto p_{\mathrm{cf}}(\mathbf{x}^{(1:N)}|\mathbf{X}^{(1:N)}, \boldsymbol{\theta}_{\mathrm{cf}}) p_{\mathrm{c}}(\mathbf{X}^{(1:N)}|\boldsymbol{\theta}_{\mathrm{c}})$.

Therefore,

$$\Rightarrow KL(q(\mathbf{X}^{(1:N)})||p(\mathbf{X}^{(1:N)}|\mathbf{x}^{(1:N)}, \boldsymbol{\theta}_{\mathrm{c}}, \boldsymbol{\theta}_{\mathrm{cf}})) = 0.$$

Expectation maximization

1. Initial parameter setting for iteration $t = 0$: $\boldsymbol{\theta}^0_{\mathrm{cf}}$ and $\boldsymbol{\theta}^0_{\mathrm{c}}$
2. E-step: $q^{(t+1)} = \arg\max_q \mathcal{L}(q, \boldsymbol{\theta}^t)$
   Involves MCMC to obtain samples $\mathbf{X} \sim q(\mathbf{X}^{(1:N)}|\mathbf{x}^{(1:N)}, \boldsymbol{\theta}^t_{\mathrm{cf}}, \boldsymbol{\theta}^t_{\mathrm{cf}})$
3. M-step: $\boldsymbol{\theta}^{t+1} = \arg\max_{\boldsymbol{\theta}} \mathcal{L}(q^{t+1}, \boldsymbol{\theta}^t)$
4. $t \to t + 1$

Derivatives

First order derivatives

$$\frac{\partial \mathcal{L}}{\partial \boldsymbol{\theta}_\mathrm{c}} = \beta \left(N \left\langle \frac{\partial U_\mathrm{c}(\mathbf{X}, \boldsymbol{\theta}_\mathrm{c})}{\partial \boldsymbol{\theta}_\mathrm{c}} \right\rangle_{p_\mathrm{c}(\mathbf{x}^{(i)}|\boldsymbol{\theta}_\mathrm{c}^t)} - \sum_{i=1}^{N} \left\langle \frac{\partial U_\mathrm{c}(\mathbf{X}^{(i)}, \boldsymbol{\theta}_\mathrm{c})}{\partial \boldsymbol{\theta}_\mathrm{c}} \right\rangle_{q(\mathbf{X}^{(i)}|\mathbf{x}^{(i)}, \boldsymbol{\theta}_\mathrm{cf}^t, \boldsymbol{\theta}_\mathrm{c}^t)} \right)$$

$$\frac{\partial \mathcal{L}}{\partial \boldsymbol{\theta}_\mathrm{cf}} = \left\langle \frac{1}{p_\mathrm{cf}(\mathbf{x}^{(1:N)}|\mathbf{X}, \boldsymbol{\theta}_\mathrm{cf})} \frac{\partial p_\mathrm{cf}(\mathbf{x}^{(1:N)}|\mathbf{X}, \boldsymbol{\theta}_\mathrm{cf})}{\partial \boldsymbol{\theta}_\mathrm{cf}} \right\rangle_{q(\mathbf{X}^{(1:N)}|\mathbf{x}^{(1:N)}, \boldsymbol{\theta}_\mathrm{cf}^t, \boldsymbol{\theta}_\mathrm{c}^t)}$$

Second order derivatives

$$\begin{aligned}
\frac{\partial^2 \mathcal{L}}{\partial \theta_k \partial \theta_l} = &- \beta \sum_{i=1}^{N} \left\langle \frac{\partial^2 U_\mathrm{c}(\mathbf{X}^{(i)}, \boldsymbol{\theta}_\mathrm{c})}{\partial \theta_k \partial \theta_l} \right\rangle_{q(\mathbf{X}^{(i)}|\mathbf{x}^{(i)}, \boldsymbol{\theta}'_\mathrm{cf}, \boldsymbol{\theta}'_\mathrm{c})} + \beta N \left\langle \frac{\partial^2 U_\mathrm{c}(\mathbf{X}, \boldsymbol{\theta}_\mathrm{c})}{\partial \theta_k \partial \theta_l} \right\rangle_{p_\mathrm{c}(\mathbf{X}|\boldsymbol{\theta}_\mathrm{c})} \\
&- \beta^2 N \left\langle \frac{\partial U_\mathrm{c}(\mathbf{X}, \boldsymbol{\theta}_\mathrm{c})}{\partial \theta_k} \frac{\partial U_\mathrm{c}(\mathbf{X}, \boldsymbol{\theta}_\mathrm{c})}{\partial \theta_l} \right\rangle_{p_\mathrm{c}(\mathbf{X}|\boldsymbol{\theta}_\mathrm{c})} \\
&+ \beta^2 N \left\langle \frac{\partial U_\mathrm{c}(\mathbf{X}, \boldsymbol{\theta}_\mathrm{c})}{\partial \theta_k} \right\rangle_{p_\mathrm{c}(\mathbf{X}|\boldsymbol{\theta}_\mathrm{c})} \left\langle \frac{\partial U_\mathrm{c}(\mathbf{X}, \boldsymbol{\theta}_\mathrm{c})}{\partial \theta_l} \right\rangle_{p_\mathrm{c}(\mathbf{X}|\boldsymbol{\theta}_\mathrm{c})}
\end{aligned}$$

Derivatives

Second order derivatives

$$\frac{\partial^2 \mathcal{L}(q(\mathbf{X}^{(1:N)}), \boldsymbol{\theta}_\mathrm{c}, \boldsymbol{\theta}_\mathrm{cf})}{\partial \boldsymbol{\theta}_\mathrm{cf}^2} = \sum_{i=1}^{N} \left\langle \frac{1}{p_\mathrm{cf}(\mathbf{x}^{(i)}|\mathbf{X}^{(i)}, \boldsymbol{\theta}_\mathrm{cf})} \frac{\partial^2 p_\mathrm{cf}(\mathbf{x}^{(i)}|\mathbf{X}^{(i)}, \boldsymbol{\theta}_\mathrm{cf})}{\partial \boldsymbol{\theta}_\mathrm{cf}^2} - \frac{1}{p_\mathrm{cf}(\mathbf{x}^{(i)}|\mathbf{X}^{(i)}, \boldsymbol{\theta}_\mathrm{cf})^2} \left(\frac{\partial p_\mathrm{cf}(\mathbf{x}^{(i)}|\mathbf{X}^{(i)}, \boldsymbol{\theta}_\mathrm{cf})}{\partial \boldsymbol{\theta}_\mathrm{cf}} \right)$$

Robbins-Monro Optimization

Gradients are sample averages $\rightarrow$ noise afflicted.

Robbins-Monro Algorithm

$$\boldsymbol{\theta}^{t+1} = \boldsymbol{\theta}^{t} + \alpha_t \nabla_{\boldsymbol{\theta}} \mathcal{L}(\boldsymbol{\theta}^{t})$$

with

$$\alpha_t = \frac{\alpha}{(t+A)^{\rho}}, \text{ with } \frac{1}{2} < \rho \leq 1.$$

- Convergence is guaranteed to the true optimum $\boldsymbol{\theta}^*$ if infinite steps performed.
- α_t is a sequence of real numbers and has to fulfill:

$$\sum_{t=1}^{\infty} \alpha_t = \infty \text{ and } \sum_{t=1}^{\infty} \alpha_t^2 < \infty$$

Laplace Approximation

Approximating a density function $p(\mathbf{z}) = \frac{1}{Z} f(\mathbf{z})$ by

$$q(\mathbf{z}) = \mathcal{N}(\mathbf{z}|\mathbf{z}_0, \mathbf{\Sigma})$$

with $\mathbf{\Sigma} = \left(-\nabla\nabla f(\mathbf{z})|_{\mathbf{z}=0}\right)^{-1}$ and $\nabla f(\mathbf{z})|_{\mathbf{z}=\mathbf{z}_0} = 0$.

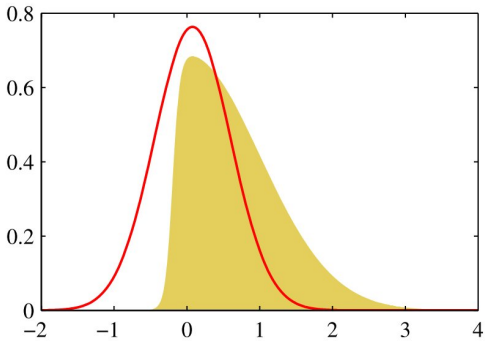


Figure : Laplace approximation[3]

For the given problem using a noniformative prior it follows

Laplace approximation

$$\frac{\partial^2 \log p(\boldsymbol{\theta}_{\text{cf}}|\mathbf{x}^{(1:N)})}{\partial \boldsymbol{\theta}_{\text{cf}}^2} = \sum_i^N \left\langle \frac{\partial^2 \log p_{\text{cf}}(\mathbf{x}^{(i)}|\mathbf{X}^{(i)}, \boldsymbol{\theta}_{\text{cf}})}{\partial \boldsymbol{\theta}_{\text{cf}}^2} \right\rangle_{q(\mathbf{x}^{(i)})}$$

Predictive uncertainty

Propagate uncertainty induced by mapping

1. $\boldsymbol{\theta}_{\text{cf}}^{i} \sim \mathcal{N}(\boldsymbol{\theta}_{\text{cf}}^{\text{MAP}}, \boldsymbol{\Sigma})$ with $\boldsymbol{\Sigma} = \left(- \frac{\partial^2 \log p(\boldsymbol{\theta}_{\text{cf}} | \mathbf{x}^{(1:N)})}{\partial \boldsymbol{\theta}_{\text{cf}}^2} \right)^{-1}$
2. $\mathbf{X}^{ij} \sim p_{\text{c}}(\mathbf{X} | \boldsymbol{\theta}_{\text{c}}^{\text{MAP}})$
3. $\mathbf{x}_k^{ij} \sim p_{\text{cf}}(\mathbf{x} | \mathbf{X}^{ij}, \boldsymbol{\theta}_{\text{cf}}^{i})$

Predictice uncertainty of properties:

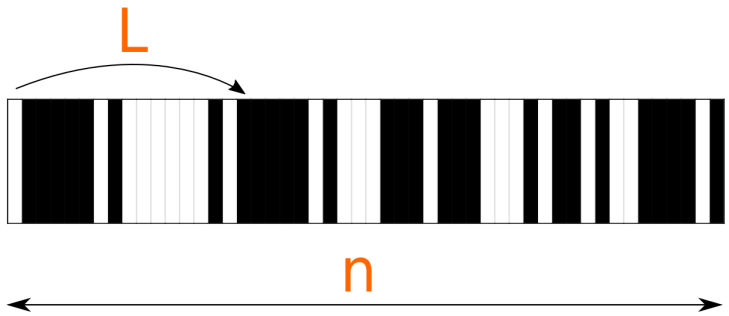
$$\boxed{\left\langle f|_{\boldsymbol{\theta}_{\text{c}}, \boldsymbol{\theta}_{\text{cf}}^{i}} = \right\rangle = \int f(\mathbf{x}) p_{\text{cf}}(\mathbf{x} | \mathbf{X}, \boldsymbol{\theta}_{\text{cf}}^{i}) p_{\text{c}}(\mathbf{X} | \boldsymbol{\theta}_{\text{c}}^{\text{MAP}}) d\mathbf{X} d\mathbf{x}}$$

Example Problem: Coarse-graining 1D Ising Model

- Applying proposed method ('predCg') for coarse-graining a one dimensional Ising Model
- Comparison between *'predCg'* and the deterministic formulation of minimizing the KL-divergence ('relEntr') by Shell (2008)
- Assessed by predictive capabilities for magnetization and correlation

Ising Model - System Setting

The Ising-model regarded in the following discussion is one dimensional. It has the properties:



- System size n_f in fine- and n_c in coarse-scale
- Level of coarse-graining $l_c = \frac{n_f}{n_c}$
- Inverse temperature β
- External field μ
- Regarded interaction length L_f in fine- and L_c in coarse scale, including all interactions $k \leq L$
- $J_0 = 1$ overall interaction strength

Ising Model - Potential in Fine-Scale

Fine variables take the values $x_i \in \{-1, 1\}$ following $p_f(\mathbf{x}|\beta) \propto \exp(-\beta U(\mathbf{x}, J_k, \mu))$:

Fine-scale potential

$$U(\mathbf{x}, J_k, \mu) = -\frac{1}{2} \sum_{k=1}^{L_f} J_k \sum_{|i-j|=k} x_i x_j - \mu \sum_{i=1}^{n_f} x_i$$

with $i, j \in \{1, \ldots, n_f\}$ having n_f lattice sites.

- Maximal interactions of L_f sites apart are regarded in the potential.
- $|i - j| = k$ interpreted as summation over neighbors k over all sites i sites apart
- J_k, strength of the k-th interaction.

with J_k following a power law for a given overall strength J_0 and exponent a, $J_k = \frac{K}{Lk^a}$ with,

$$K = J_0 L^{1-a} \sum_{k=1}^{L} k^{-a}$$

in order to normalize the interaction strength [1].

Ising Model - Potential in Coarse-Scale

Coarse variables take the values $X_i \in \{-1, 1\}$, $p_c \propto \exp\{-\beta U_c(\mathbf{X}, \boldsymbol{\theta}_c, \mu)\}$

Coarse-scale potential

$$
\begin{aligned}
U_c(\mathbf{X}, \boldsymbol{\theta}_c, \mu) = -\frac{1}{2}(&\theta^{lin} \sum_{i=1}^{n_c} X_i \\
&+ \sum_{k=1}^{L_c} \theta_k^{twop} \sum_{|i-j|=k} X_i X_j \\
&+ \sum_{i=1}^{n_c} \theta_{mn}^{trip} \sum_{m=1} \sum_{n=1} x_i x_{i\pm m} x_{i\pm m\pm n}) \\
&- \mu \sum_{i=1}^{n_c} X_i
\end{aligned}
$$

with $i, j \in \{1, \ldots, n_c\}$ and $n_c \ll n_f$ lattice sites.

Ising Model - Mapping Fine to Coarse (for relative entropy)

Each coarse variable X_r, $r = \{1, \ldots, R\}$ describes S fine-scale variables $x_{r,s}$ with $s = \{1, \ldots, S\}$. The mapping from fine-scale variables $x_{r,s}$ to coarse variable X_r is defined as,

Fine-to-coarse mapping

$$X_r = \begin{cases} +1, & \frac{1}{S}\sum_s^S x_{r,s} > 0 \\ -1, & \frac{1}{S}\sum_s^S x_{r,s} < 0 \\ U(-1,+1), & \text{otherwise} \end{cases}.$$

Figure : Mapping from fine-scale **x** to coarse-scale **X**

Ising Model - Mapping Coarse to Fine

Coarse-to-fine mapping $p_{\mathrm{cf}}(\mathbf{x}|\mathbf{X}, \theta_{\mathrm{cf}})$

$$p_{\mathrm{cf}}(x_{r,s}|X_r) = \theta_{\mathrm{cf}}^{\frac{1+x_{r,s}X_r}{2}} (1-\theta_{\mathrm{cf}})^{\frac{1+x_{r,s}X_r}{2}}$$

Assumption: x_i conditionally independent for given $\mathbf{X}$:

$$\begin{aligned} p(\mathbf{x}^{(1:N)}|\mathbf{X}^{(1:N)}, \boldsymbol{\theta}_{\mathrm{cf}}) &= \prod_{i=1}^{N} p_{\mathrm{cf}}(\mathbf{x}^{(t)}|\mathbf{X}^{(t)}, \boldsymbol{\theta}_{\mathrm{cf}}) \\ &= \prod_{i=1}^{N}\prod_{r=1}^{R}\prod_{s=1}^{S} \left(\theta_{\mathrm{cf}}^{\frac{1+x_{r,s}^i X_r^i}{2}} (1-\theta_{\mathrm{cf}})^{\frac{1-x_{r,s}^i X_r^i}{2}} \right) \\ &= \theta_{\mathrm{cf}}^{\sum_{i=1}^{N}\sum_{r=1}^{R}\sum_{s=1}^{S} \frac{1+x_{r,s}^i X_r^i}{2}} \\ &\quad (1-\theta_{\mathrm{cf}})^{\sum_{i=1}^{N}\sum_{r=1}^{R}\sum_{s=1}^{S} \frac{1-x_{r,s}^i X_r^i}{2}} \end{aligned}$$

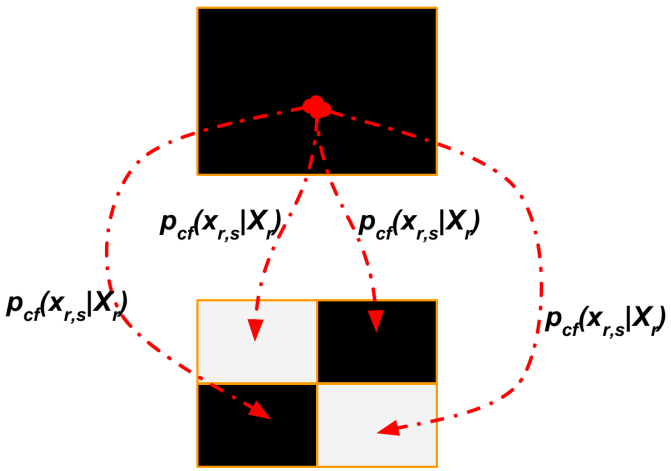


Figure : Probabilistic mapping form coarse to fine: $p_{\mathrm{cf}}(\mathbf{x}|\mathbf{X}, \boldsymbol{\theta}_{\mathrm{cf}})$

Which properties to predict?

The magnetization m is given:

Magnetization in the fine-scale

$$m_{fine} = \int \frac{1}{n} \sum_{i}^{n} x_i \; p_{\mathrm{f}}(\mathbf{x}) d\mathbf{x}$$

Magnetization calculated with proposed method ('*predcg*')

$$m_{ML,fine} = \int \frac{1}{n} \sum_{i}^{n} x_i \; p_{\mathrm{cf}}(\mathbf{x}|\mathbf{X}, \boldsymbol{\theta}_{\mathrm{cf}}) p_{\mathrm{c}}(\mathbf{X}|\boldsymbol{\theta}_{\mathrm{c}}) p(\boldsymbol{\theta}_{\mathrm{cf}}, \boldsymbol{\theta}_{\mathrm{c}}|\mathbf{x}^{(1:N)}) \;\; d\boldsymbol{\theta}_{\mathrm{cf}} d\boldsymbol{\theta}_{\mathrm{c}} d\mathbf{X} d\mathbf{x}$$

Quantity of Interest: Magnetization

Magnetization calculated with relative entropy method '*relEntr*'

(deterministic mapping fine to coarse: $\mathbf{X} = \boldsymbol{\xi}(\mathbf{x})$)

- For each μ calculate optimal model parameter $\boldsymbol{\theta}^*_{\text{cg, det}}$
- Since we want to compare the magnetization on the same (fine) scale a mapping is introduced
  - if $X_r = -1$, select randomly a possible configuration $x_{r,s}$ where either all $x_{r,s} = -1$ or one is $x_{r,s} = -1$ and the other $x_{r,s} = 1$ (for given r: $s \in \{1 \dots l_c\}$)
  - if $X_r = +1$, select randomly a possible configuration $x_{r,s}$ where either all $x_{r,s} = +1$ or one is $x_{r,s} = +1$ and the other $x_{r,s} = -1$ (for given r: $s \in 1, 2$)
- For a given coarse-state the selected fine-states $\mathbf{x}_{\text{det}}$ are used for calculating the magnetization

Relative error in magnetization,

$$\boxed{\text{err}_{\text{mag}} = \frac{\|m - m_{\text{truth}}\|}{\|m_{\text{truth}}\|}.}$$

Quantity of Interest: Correlation

The *correlation* R_l of all sites being separated by l sites is measured as,

$$R_l = \frac{1}{n} \sum_{|i-j|=l}^{n} < x_i x_j > .$$

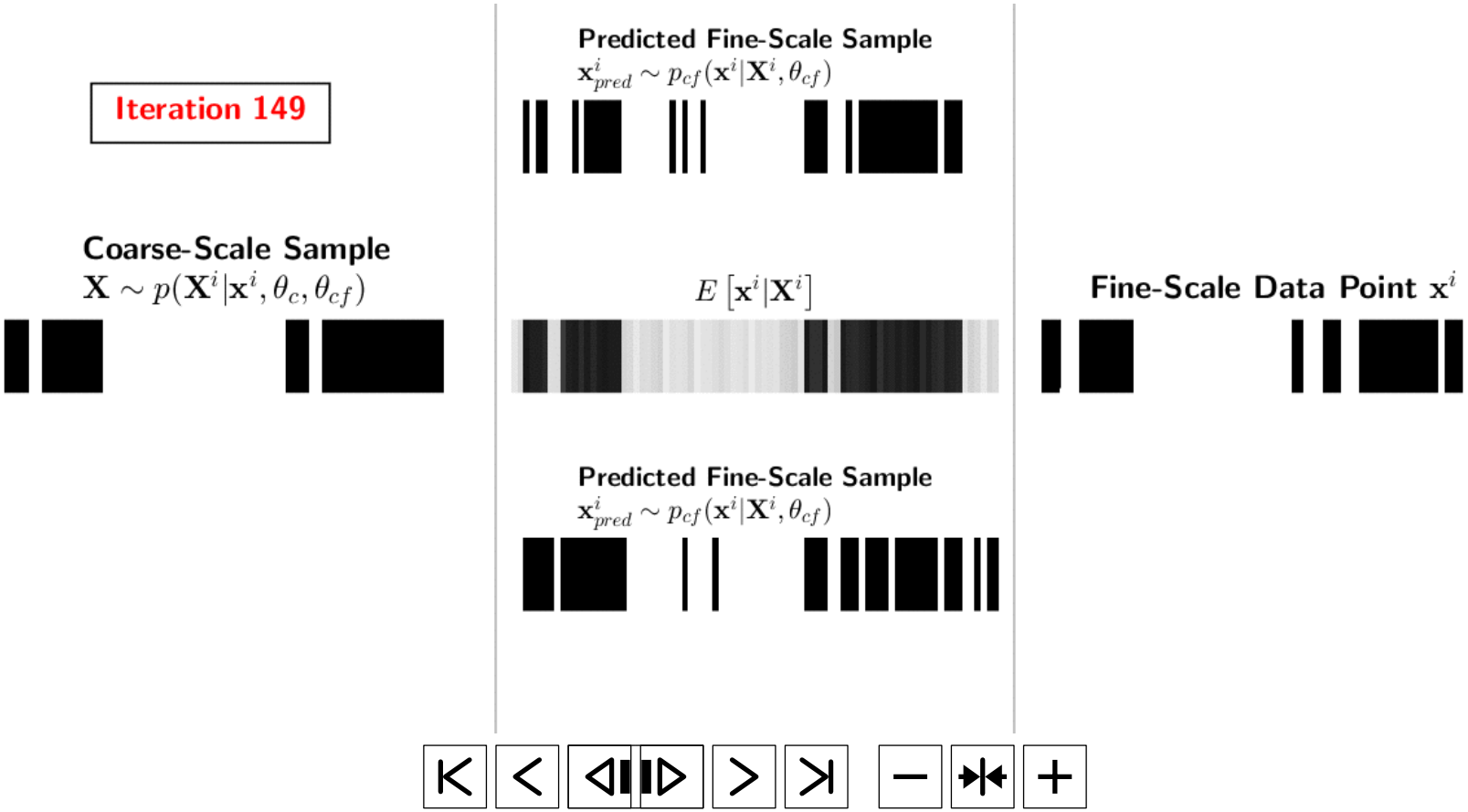
Relative correlation error,

$$\boxed{\text{err}_{\text{corr}} = \frac{\| R - R_{\text{truth}} \|}{\| R_{\text{truth}} \|}.}$$

Ising Model - ML Algorithm

for $\mu = \mu_{min}$ to μ_{max} **do**
  set initial fine-scale parameter: J_0, n_f, μ
  create data set: $\mathbf{x}^{(1:S_f)} \sim p(\mathbf{x}^{(1:S_f)} | J_0, \mu, \beta)$ with S_f samples
  set initial coarse-graining parameter: $\boldsymbol{\theta}_{c0}$, $\boldsymbol{\theta}_{cf0}$, $n_c = \frac{n_f}{l_c}$, μ
  $\boldsymbol{\theta}_c^t \leftarrow \boldsymbol{\theta}_{c0}$
  $\boldsymbol{\theta}_{cf}^t \leftarrow \boldsymbol{\theta}_{cf0}$
  while !(Convergence check passed) **do**
    function E-STEP
    $\mathbf{X} \sim q(\mathbf{X}|\mathbf{x}^{(1:N)}, \boldsymbol{\theta}_c^t, \boldsymbol{\theta}_{cf}^t, \mu) \propto p_{cf}(\mathbf{x}^{(1:N)}|\mathbf{X}, \boldsymbol{\theta}_{cf}^t) p_c(\mathbf{X}|\boldsymbol{\theta}_c^t, \mu)$ with S_q per data point $\mathbf{x}^{(i)}$
    function M-STEP
    $$\frac{\partial \mathcal{L}}{\partial \boldsymbol{\theta}_c} = -\beta \left\langle \frac{\partial U_c(\mathbf{X}, \boldsymbol{\theta}_c)}{\partial \boldsymbol{\theta}_c} \right\rangle_{q(\mathbf{X}|\mathbf{x}^{(1:N)}, \boldsymbol{\theta}_{cf}^t, \boldsymbol{\theta}_c^t)} + \beta \left\langle \frac{\partial U_c(\mathbf{X}, \boldsymbol{\theta}_c)}{\partial \boldsymbol{\theta}_c} \right\rangle_{p_c(\mathbf{X}|\boldsymbol{\theta}_c^t)}$$
    $$\frac{\partial \mathcal{L}}{\partial \boldsymbol{\theta}_{cf}} = \left\langle \frac{1}{p_{cf}(\mathbf{x}^{(1:N)}|\mathbf{X}, \boldsymbol{\theta}_{cf})} \frac{\partial p_{cf}(\mathbf{x}^{(1:N)}|\mathbf{X}, \boldsymbol{\theta}_{cf})}{\partial \boldsymbol{\theta}_{cf}} \right\rangle_{q(\mathbf{X}|\mathbf{x}, \boldsymbol{\theta}_{cf}^t, \boldsymbol{\theta}_c^t)}$$
    $\boldsymbol{\theta}_c^{t+1} \leftarrow \boldsymbol{\theta}_c^t + \alpha_{\boldsymbol{\theta}_c} \frac{\partial \mathcal{L}}{\partial \boldsymbol{\theta}_c}$
    $\boldsymbol{\theta}_{cf}^{t+1} \leftarrow \boldsymbol{\theta}_{cf}^t + \alpha_{\boldsymbol{\theta}_{cf}} \frac{\partial \mathcal{L}}{\partial \boldsymbol{\theta}_{cf}}$
  end while
  $\mathbf{x}_{pred} \sim \tilde{p}(\mathbf{x}_{new}|\boldsymbol{\theta}_c^t, \boldsymbol{\theta}_{cf}^t)$ with $S_{pred} * S_c$ samples
  function EVALMAG($\mathbf{x}_{pred}$, $\mathbf{x}^{(1:N)}$)
  function EVALCORR($\mathbf{x}_{pred}$, $\mathbf{x}^{(1:N)}$)
end for=0

Ising Model - Posterior and Prediction



Overview

1. Behavior of parameters with respect to μ
2. Influence of available data S_f for training the model
3. Level of coarse-graining
4. Aspects of coarse-grained models

Behavior of parameters with respect to μ

Constant attributes:

- $S_c = 200$ (samples $\mathbf{X} \sim p_c(\mathbf{X}|\theta_c)$)
- $S_q = 50$ (samples $\mathbf{X} \sim q(\mathbf{X}^{(i)}|\mathbf{x}^{(i)}, \theta_c, \theta_{cf})$)
- $J_0 = 1$ (overall interaction strength)
- $S_{pred} = 100 S_c$ (samples for prediction step after learning the model)

Furthermore:

- $n_f = 32$, system size fine-scale
- $L_f = L_c = 1$, in fine- and coarse scale nearest-neighbor interactions are regarded

Behavior of parametrization with respect to μ

The potentials follows to,

$$U(J,\mu) = -\frac{1}{2}J \sum_{|i-j|=1} x_i x_j - \mu \sum_{i=1}^{n_\mathrm{f}} x_i$$

$$U_\mathrm{c}(\boldsymbol{\theta}_\mathrm{c},\mu) = -\frac{1}{2}\theta_\mathrm{c} \sum_{|i-j|=1} X_i X_j - \mu \sum_{i=1}^{n_\mathrm{c}} X_i.$$

- Learn two parameters $\boldsymbol{\theta}_\mathrm{c}$ and $\boldsymbol{\theta}_\mathrm{cf}$ of the mapping p_cf
- Learning for every evaluated μ: $\mu = [-4.5, 4.5]$ with a step size of 0.6

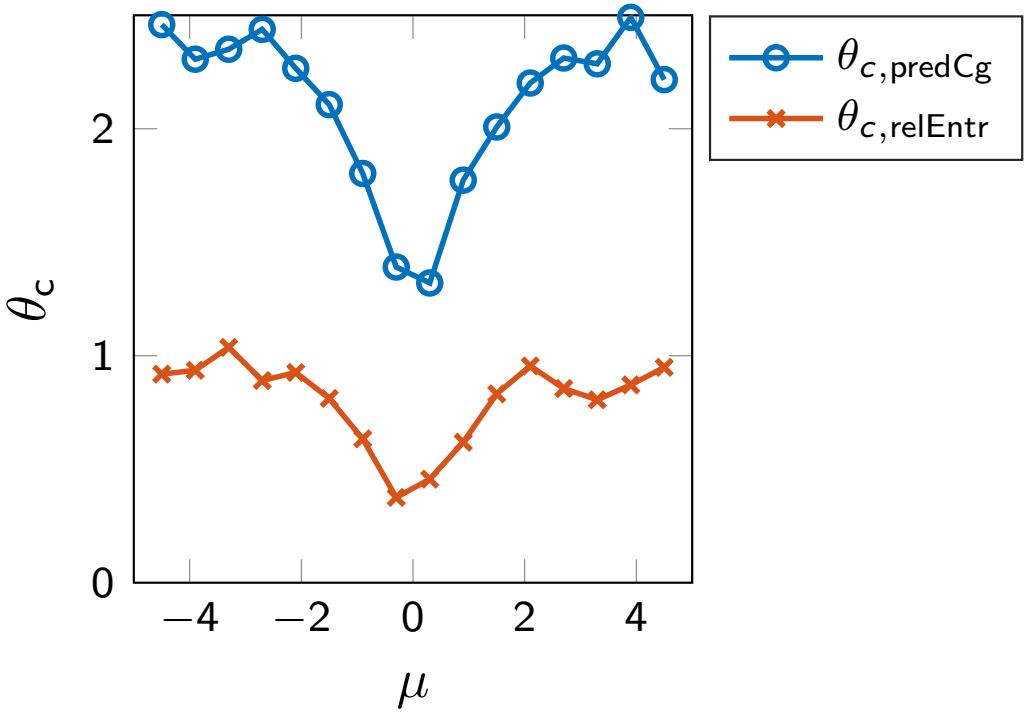


Figure : θ_c: Parameter of potential U_c

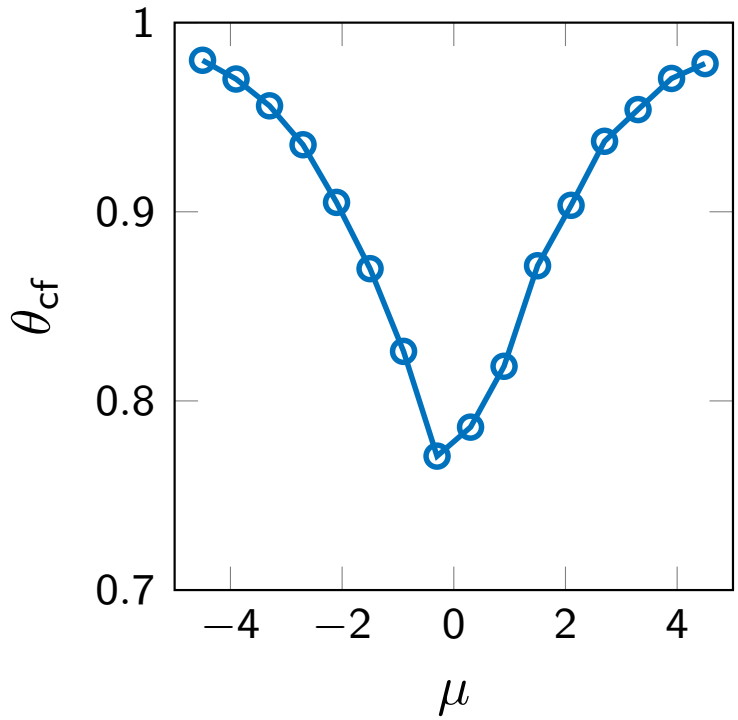


Figure : θ_{cf}: Parameter of the mapping $p_{cf}(\mathbf{x}|\mathbf{X}, \theta_{cf})$

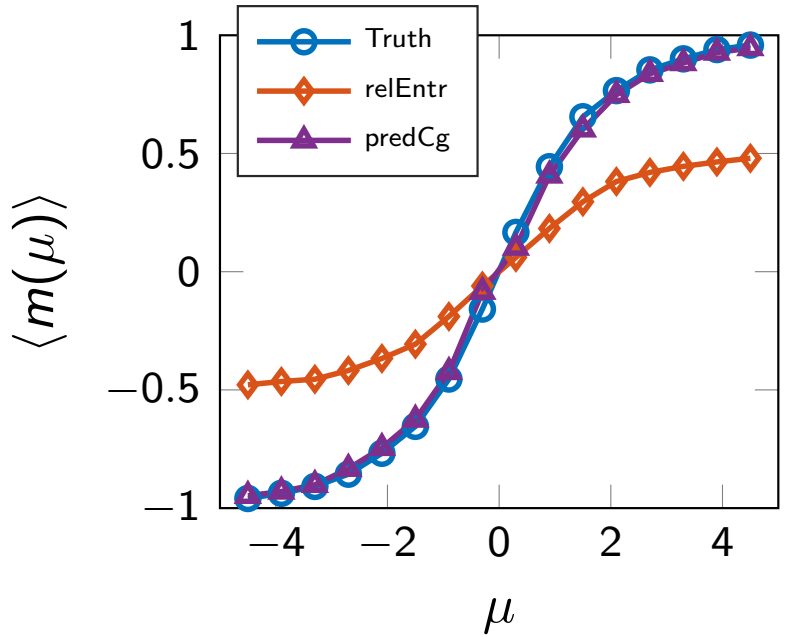


Figure : Magnetization

Key aspects

- True magnetization based on data set and predicted by 'predCg' coincides.
- Method 'relEntr' not able to predict data set due to missing mapping capabilities from coarse to fine.
- Parameters $\boldsymbol{\theta}_{\mathrm{c}}$ and $\boldsymbol{\theta}_{\mathrm{cf}}$ behave symmetrical with respect to $\mu = 0$.
- Stronger interaction $\boldsymbol{\theta}_{\mathrm{c}}$ for bigger $|\mu|$ but also higher probability $\boldsymbol{\theta}_{\mathrm{cf}}$ that the coarse configuration reflects the fine configuration.

Behavior with respect to size of data-set

- Learn $\boldsymbol{\theta}_{\mathrm{c}}$ and $\boldsymbol{\theta}_{\mathrm{cf}}$ at each evaluated μ
- Given different size of data-set S_{f} with $S_{\mathrm{f}} \in \{5, 10, 20, 50\}$.
- Interaction length in fine-scale $L_f = 10$ with exponential decay of overall strength $J_0 = 1.5$

Potentials

$$U_{\mathrm{f}}(J, \mu) = -\frac{1}{2} \sum_{k}^{L_f} J_k \sum_{|i-j|=k} x_i x_j - \mu \sum_{i=1}^{n_{\mathrm{f}}} x_i$$

$$U_{\mathrm{c}}(\boldsymbol{\theta}_{\mathrm{c}}, \mu) = -\frac{1}{2}\theta_{\mathrm{c}} \sum_{|i-j|=1} X_i X_j - \frac{1}{2}\theta_{\mathrm{c}}^{lin} \sum_{i}^{n_{\mathrm{c}}} X_i - \mu \sum_{i=1}^{n_{\mathrm{c}}} X_i.$$

Behavior with respect to size of data-set

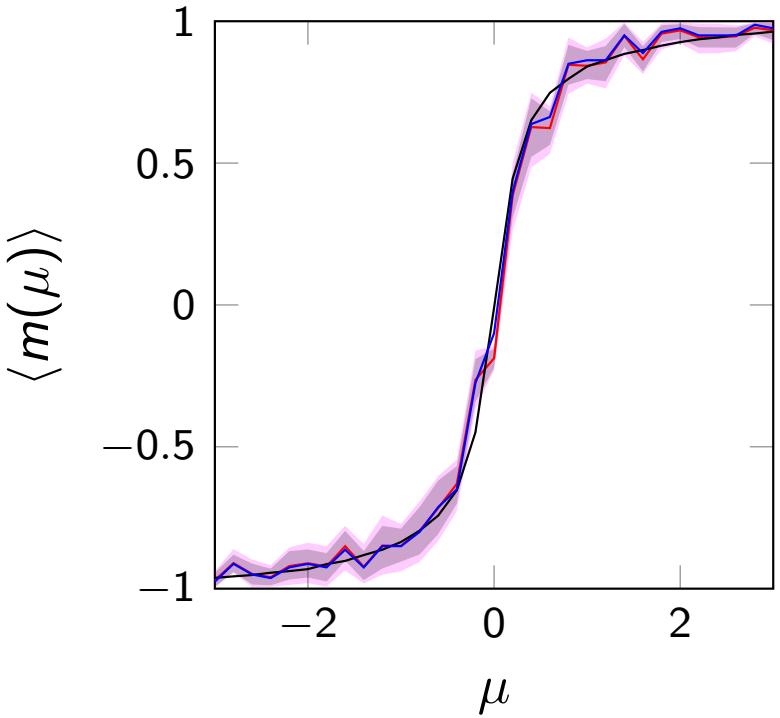
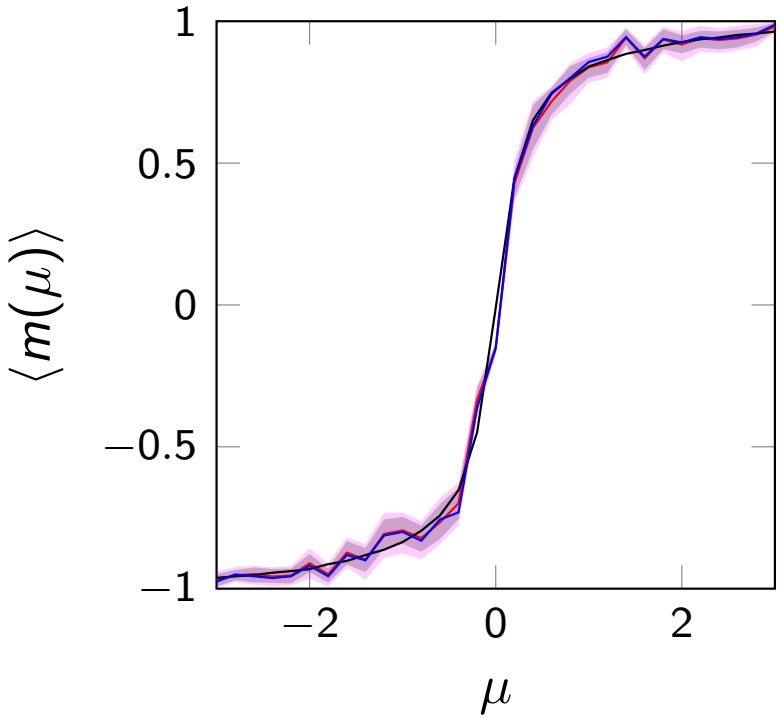


Figure : Uncertainty in magnetization $S_f = 5$



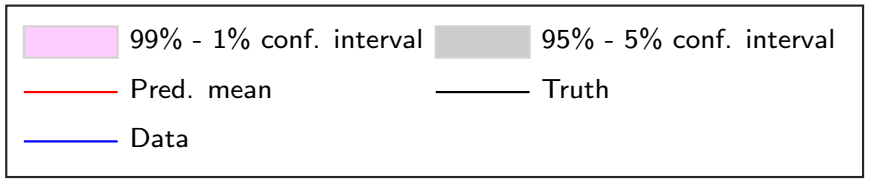


Figure : Uncertainty in magnetization $S_f = 10$

Behavior with respect to size of data-set

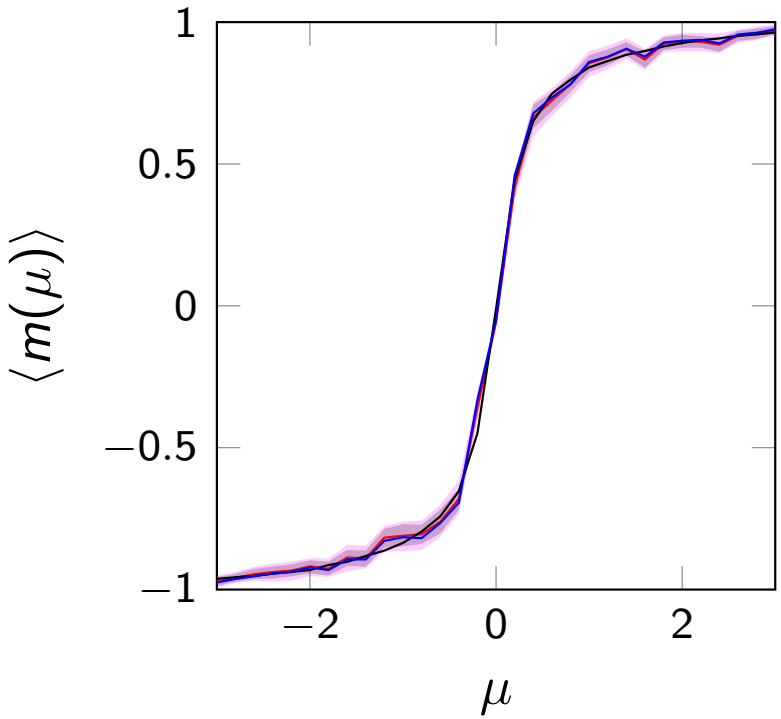


Figure : Uncertainty in magnetization $S_f = 20$

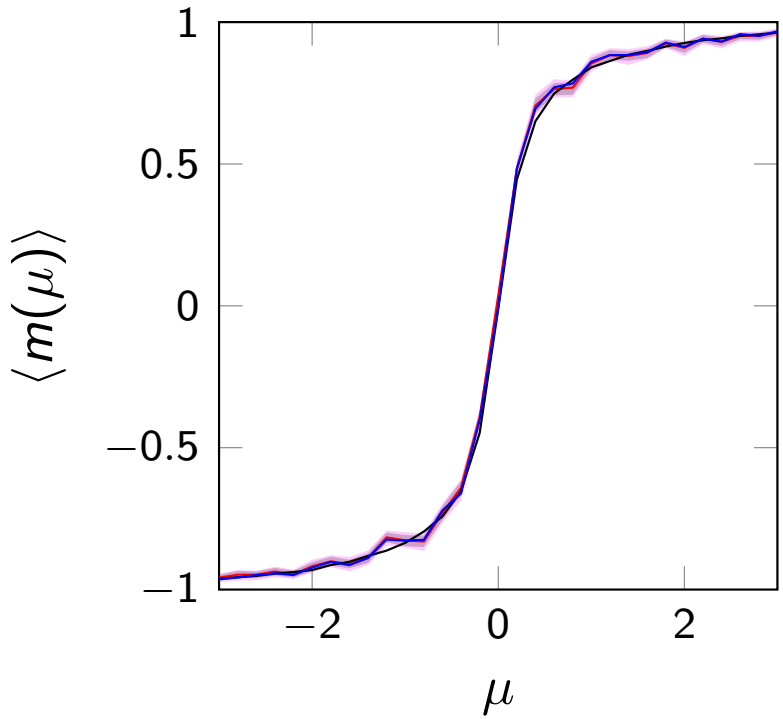


Figure : Uncertainty in magnetization $S_f = 50$

Behavior with respect to size of data-set

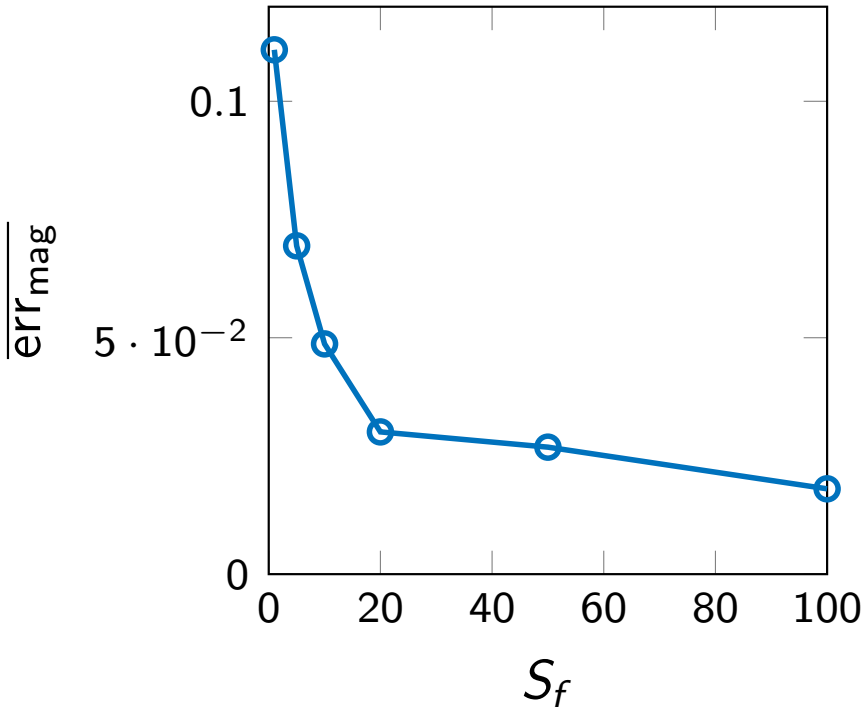


Figure : Relative mean error of magnetization with respect to increasing amount of data S_f, $S_f \in \{5, 10, 20, 50, 100\}$

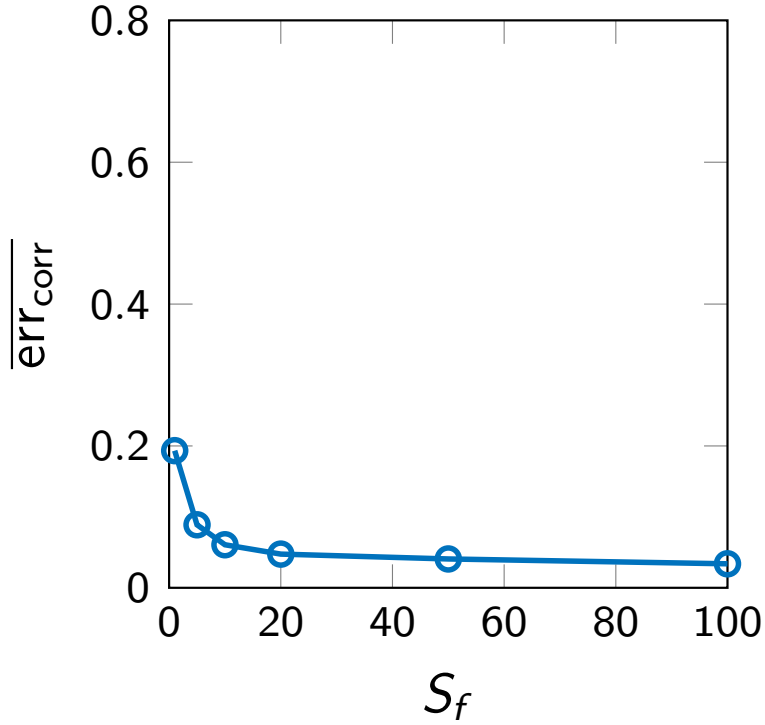


Figure : Relative mean error of correlation with respect to increasing amount of data S_f, $S_f \in \{5, 10, 20, 50, 100\}$

Behavior with respect to size of data-set

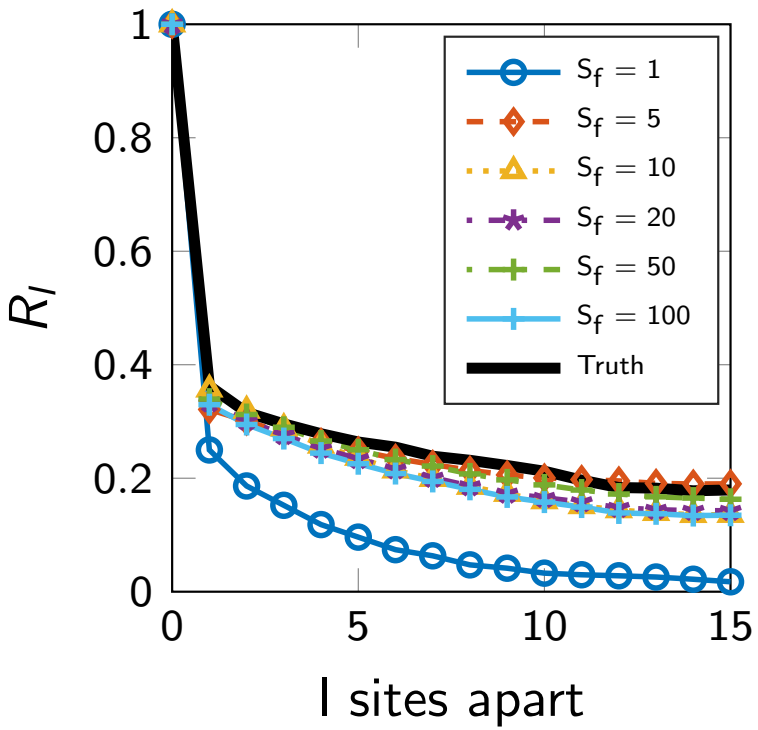
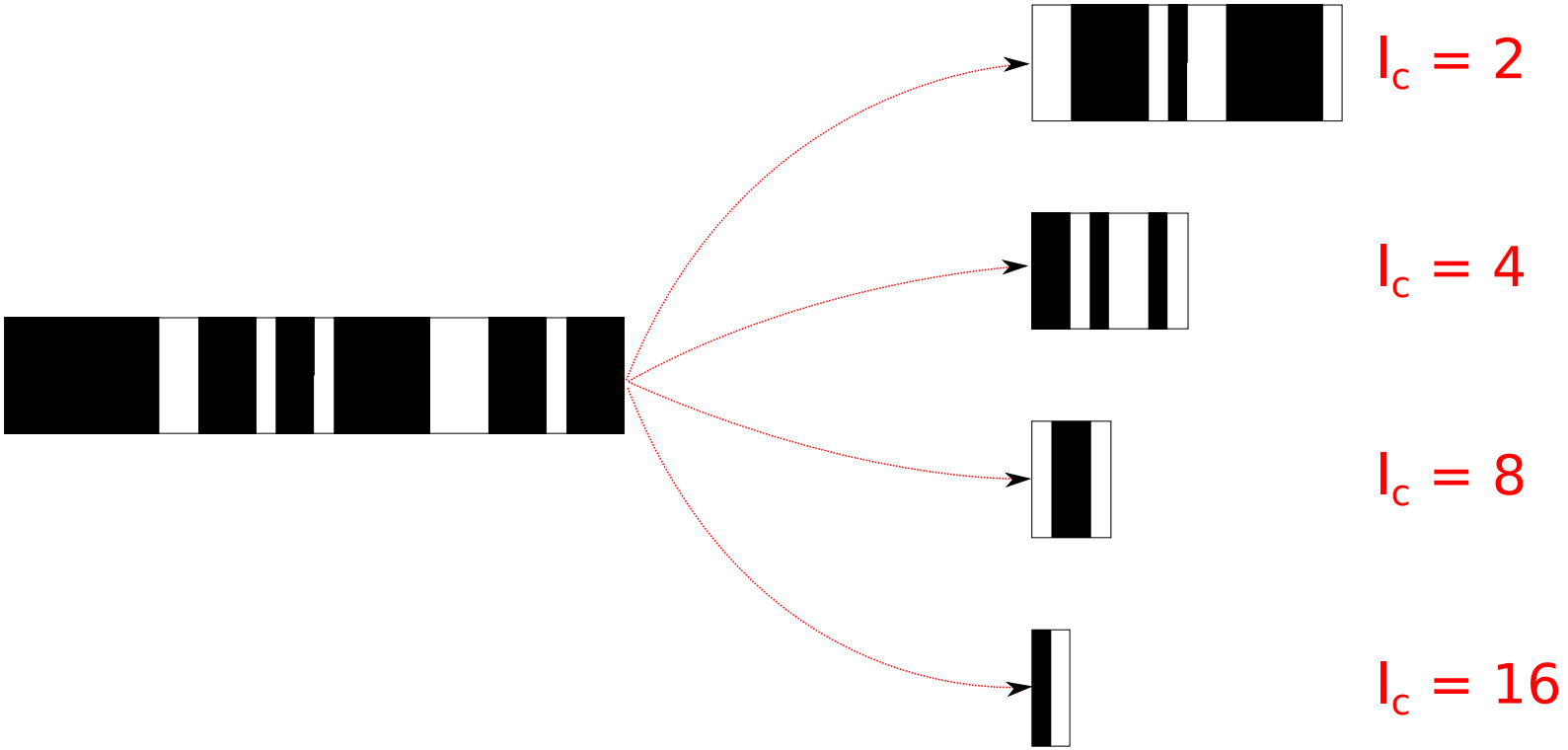


Figure : Correlation R_l for $\mu = 0$ with $S_f \in \{5, 10, 20, 50, 100\}$

- Uncertainty decreasing with increasing amount of data available
- By using $S_f = 5$ predictions resembling true magnetization well

Predictability by various levels of coarse-gaining

- Choose various level of coarse-graining l_c
- Mapping $p_{cf}(\mathbf{x}|\mathbf{X}, \theta_{cf})$ responsible for $l_c \in \{2, 4, 8, 16\}$ fine variables
- Coarse-grained potential $U_c(\mathbf{X}, \theta_c)$ as given before



Predictability by various levels of coarse gaining

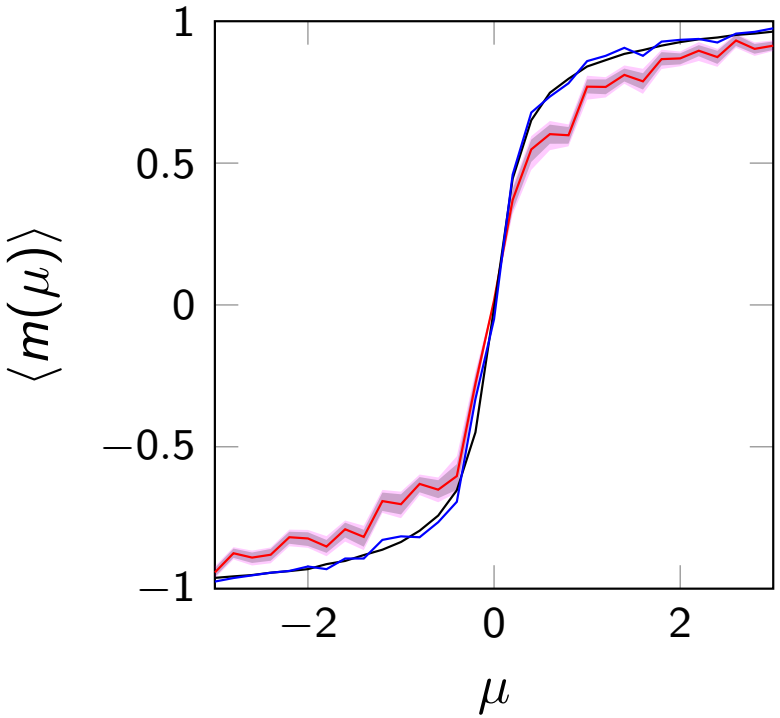
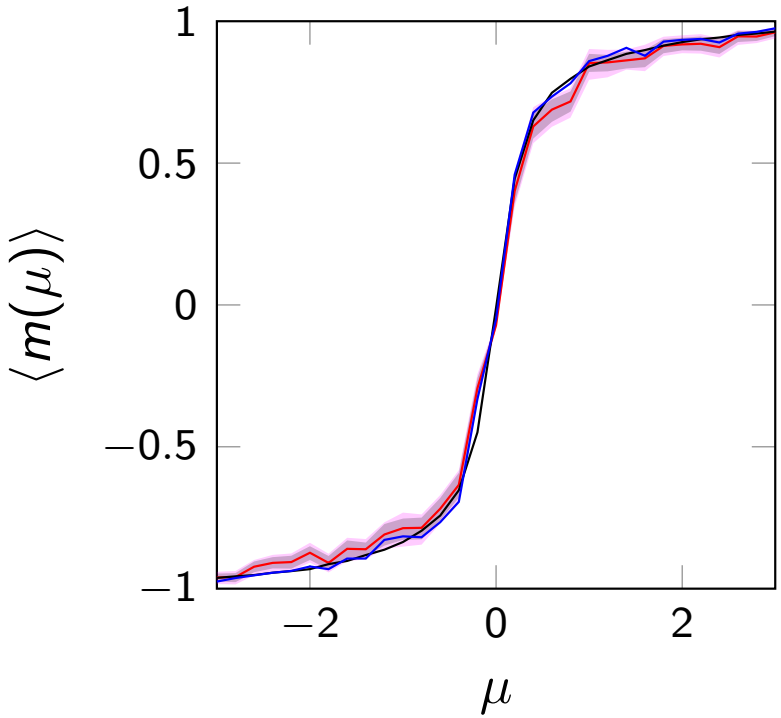


Figure : Uncertainty in magnetization $l_c = 16$



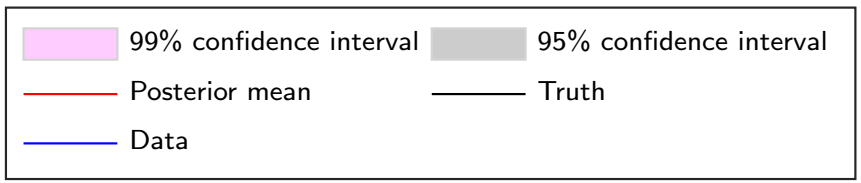


Figure : Uncertainty in magnetization $l_c = 8$

Predictability by various levels of coarse gaining

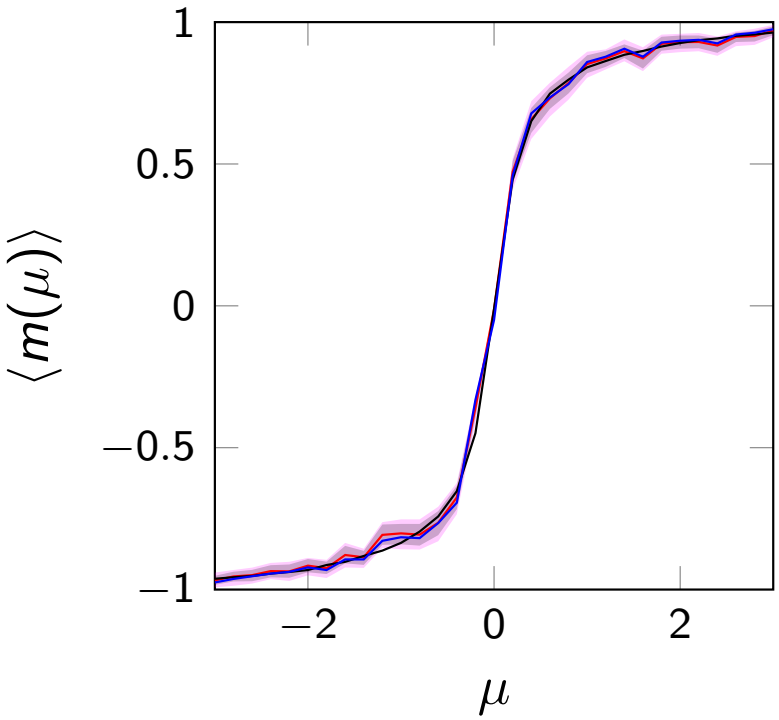


Figure : Uncertainty in magnetization $l_c = 4$

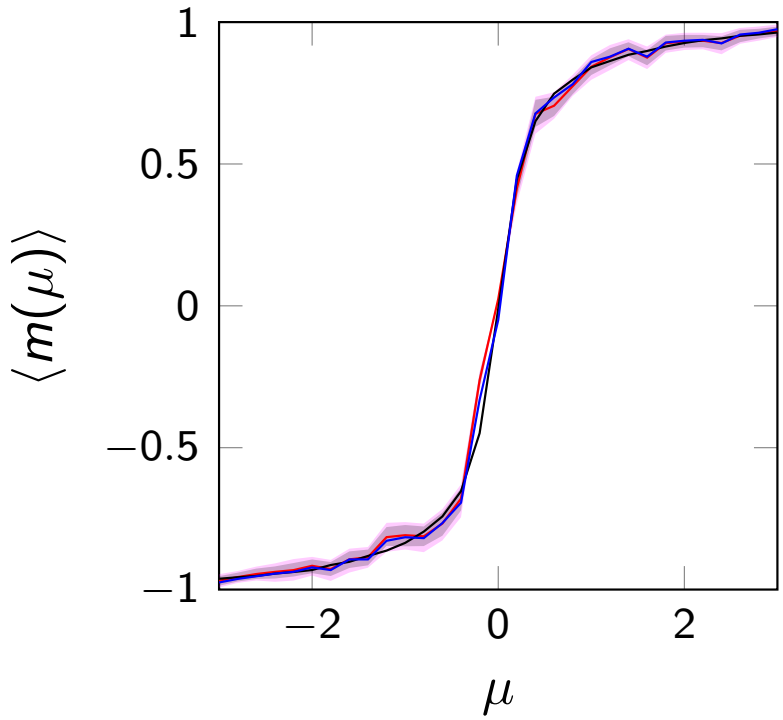


Figure : Uncertainty in magnetization $l_c = 2$

Predictability by various levels of coarse gaining

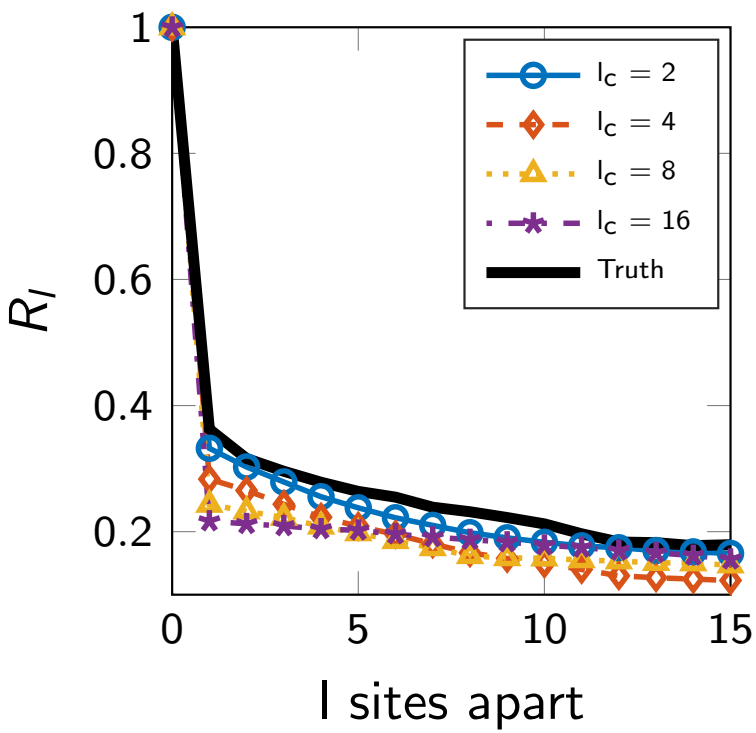
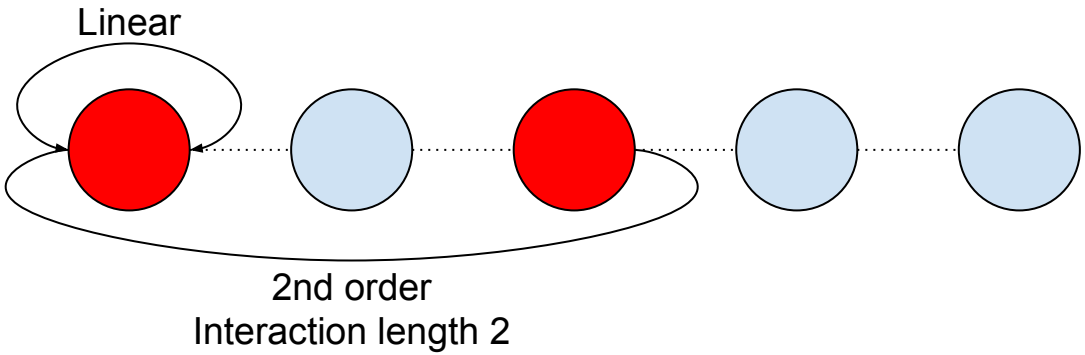


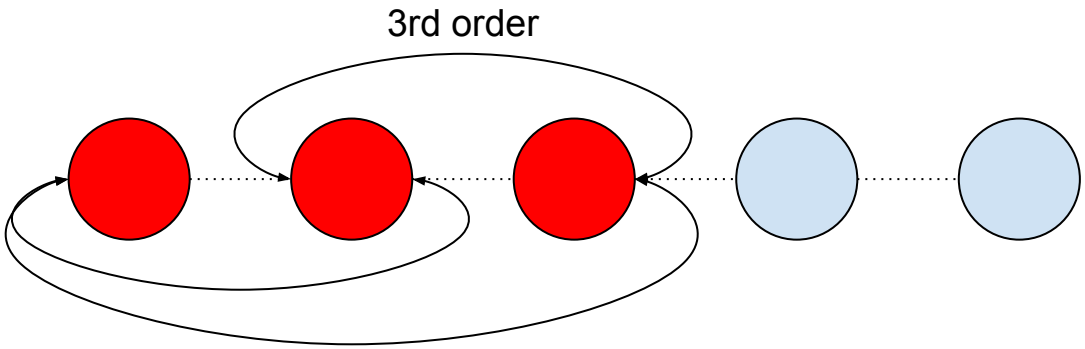
Figure : Correlation R_l at $\mu = 0$ with $l_c \in \{2, 4, 8, 16\}$

- Applying $l_c = 8$ leads already to good predictions in magnetization
- Predicting correlations needs finer resolution in coarse-scale

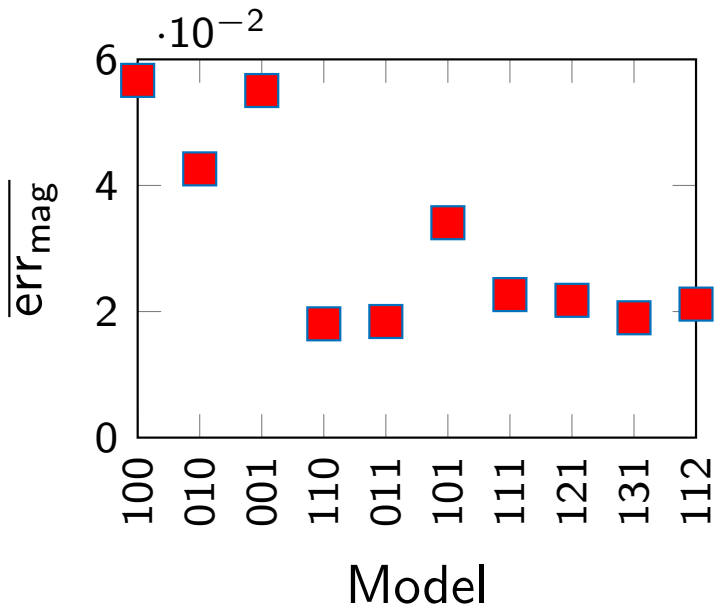
Model Comparison

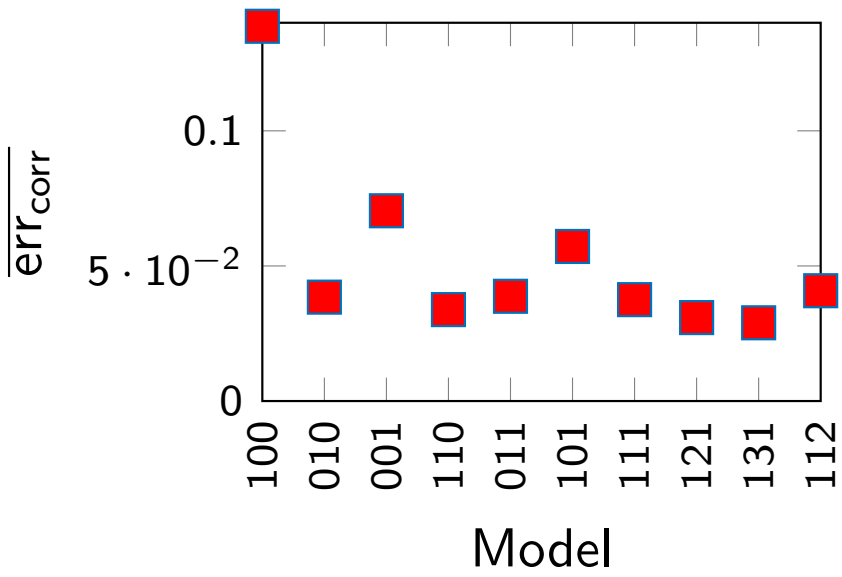
- Assumption before: Model is given for potential $U_c(\mathbf{X}, \theta_c)$
- Using different order of interactions and interaction lengths





Model Comparison





Model [abc]

- (a) $a = 0$ no linear term, $a = 1$ linear term
- (b) $b = 0$ no 2nd order term, $b = x$ 2nd order term up to interaction length $L_c = x$
- (c) $c = 0$ no third order term, $c = 1$ third order term on (nearest correlation)

Model Comparison

- Linear term necessary to describe data
- Second order interaction necessary
- In correlation: errors decreasing for including 2nd order interactions with higher interaction lengths L_c
- No further improvement for longer ranged third order interactions

Outline

Numerical issues / efficiency

- Use advanced sampling methods / optimization methods, using curvature information (in progress)
- Variational approximations

Coarse-Graining

- Hierarchical coarse-graining
- How to select mappings from coarse to fine?
- Probit- or Logit-classification model, O model for coarse variables

Algorithmic

- Approach with sparsity priors for model selection (in progress).

Bibliography I

S. Are, M. A. Katsoulakis, P. Plecháč, and L. R. Bellet.

Multibody Interactions in Coarse-Graining Schemes for Extended Systems.
SIAM Journal on Scientific Computing, 31(2):987–1015, 2008.

I. Bilionis and N. Zabaras.

A stochastic optimization approach to coarse-graining using a relative-entropy framework.
The Journal of chemical physics, 138(4):044313, Jan. 2013.

C. M. Bishop.

Pattern Recognition and Machine Learning.
Springer, 2006.

A. Chaimovich and M. S. Shell.

Coarse-graining errors and numerical optimization using a relative entropy framework.
The Journal of chemical physics, 134(9):094112, Mar. 2011.

F. Ercolessi and J. B. Adams.

Interatomic potentials from first-principles calculations: The force-matching method.
EPL (Europhysics Letters), 26(8):583, 1994.

S. Izvekov and G. A. Voth.

Multiscale coarse graining of liquid-state systems.
J. Chem. Phys., 123(13):134105, 2005.

M. Katsoulakis, P. Plecháč, and L. Rey-Bellet.

Numerical and statistical methods for the coarse-graining of many-particle stochastic systems.
Journal of Scientific Computing, 37(1):43–71, 2008.

A. P. Lyubartsev and A. Laaksonen.

Calculation of effective interaction potentials from radial distribution functions: A reverse Monte Carlo approach.
Phys. Rev. E, 52:3730–3737, Oct 1995.

Bibliography II

D. Ming and M. E. Wall.

Allostery in a coarse-grained model of protein dynamics.
Phys. Rev. Lett., 95:198103, Nov 2005.

M. S. Shell.

The relative entropy is fundamental to multiscale and inverse thermodynamic problems.
J. Chem. Phys., 129(14):144108, 2008.